Handling the reflective-formative measurement conundrum

ELSEVIER

Journal of Clinical Epidemiology 103 (2018) 71−81

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Handling the reflective-formative measurement conundrum: a practical illustration based on sustainable employability

Bram P.I. Fleuren[a,*], Ludovic G.P.M. van Amelsvoort[b], Fred R.H. Zijlstra[a], Andries de Grip[c], IJmert Kant[b]

[a]*Department of Work- and Social Psychology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands*
[b]*Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands*
[c]*Research Center for Education and the Labor Market, School of Business and Economics, Maastricht University, Maastricht, The Netherlands; IZA, Bonn, Germany; Netspar, Tilburg University, Tilburg, The Netherlands*

Accepted 11 July 2018; Published online 19 July 2018

Abstract

Objectives: Constructs capturing health or functioning can have reflective and/or formative measurement models. Although a construct's measurement model has extensive implications on the construction, validation, and use of a measurement instrument, measurement models are frequently wrongly or not explicitly specified. As this is likely due to a lack of guidelines, this study uses sustainable employability as an example to demonstrate a) the applicability of an adapted checklist for establishing a construct's measurement model; and b) the use of structural equation modeling to handle formative constructs.

Study Design and Setting: First, the checklist is applied to sustainable employability to establish its measurement model. Second, using observational self-report data from 2,544 employees, structural equation models are estimated to evaluate the structural and criterion validity of sustainable employability as a formative construct.

Results: The checklist demonstrates strong applicability, identifying sustainable employability as a formative construct. Model fit indices (Comparative fit indices $\geq$ 0.932, Tucker-Lewis indices $\geq$ 0.925, root mean square errors of approximation $\leq$ 0.034) suggest the formative measurement model for sustainable employability is valid.

Conclusion: The checklist and structural equation modeling facilitate handling formative constructs. By establishing sustainable employability as a formative construct, individuals' long term ability to function at work can be more adequately studied and intervened upon.

Keywords: Formative measurement model; Reflective measurement model; Sustainable employability; Health; Validation; Complex constructs

1. Introduction

Capturing that which cannot be directly observed is a common challenge in biomedical and social sciences. Complex multidimensional variables are abundantly constructed in an effort to capture what we think are real—or relevant—phenomena (e.g., functioning, severity of disease, life history strategies, quality of life, psychological disorders, and socioeconomic status [SES]). However, constructing such complex constructs is neither straightforward nor without pitfalls. Perhaps one of the most fundamental issues in measurement model literature is the distinction between reflective and formative measurement [1−8].

Jarvis et al. [2] describe reflective measurement models (or constructs based on effect indicators) as models ''where the covariation among the measures is caused by, and therefore reflects, variation in the underlying latent factor''. In simpler terms, the (variations in) observed item scores are, in reflective models, considered to be caused by (variations in) the underlying latent construct. For example, it is because person A's experiences a high amount of fatigue that person A scores high on items 1, 2, and 3, measuring fatigue. As the assumption of unidimensionality prescribes,

Conflict of interest: none.

* Corresponding author. Bram Fleuren Universiteitssingel 40, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Tel.: +31433884501.

E-mail address: bram.fleuren@maastrichtuniversity.nl (B.P.I. Fleuren).

https://doi.org/10.1016/j.jclinepi.2018.07.007
0895-4356/

What is new?

Key findings

- The article demonstrates that the appropriate nature of a construct's measurement model as reflective and/or formative can be identified by using an easily applicable checklist.
- The article shows that structural equation modeling can be used to evaluate the viability of modeling formative constructs and assessing such constructs' criterion validity.
- The article identifies the broad and relevant novel occupational health construct of sustainable employability as a formative construct and show that it can be effectively modeled as such.

What this adds to what was known?

- The article provides an approach to identifying and handling formative constructs, which is highly necessary as formative constructs continue to be wrongly specified and validated as if they were reflective.
- The article introduces and elaborates on the formative-reflective measurement model distinction in the field of clinical epidemiology.
- The article broadens the general understanding of the concept of sustainable employability so that research and practice can use this construct appropriately in identifying its determinants as well as designing interventions to improve it.

What is the implication and what should change now?

- The findings and reflections thereon suggest that existing constructs related to health and functioning may require careful reconsideration in terms of the appropriateness of applied measurement models, validation thereof, and structural relationships with other constructs.

all indicators (items) measure the same underlying construct, they are supposed to be correlated, and show internal consistency. Moreover, removing indicators from the measurement model does not alter the meaning of the construct as all indicators reflect the same underlying construct [2]. For example, if one of the three items measuring fatigue is removed, the remaining two items still measure fatigue and fatigue could still be measured adequately (Fig. 1A).

Formative measurement models (or constructs based on causal or composite indicators) are conceptualized as

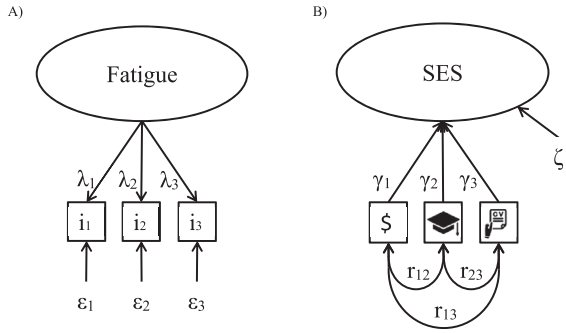

Fig. 1. Examples of a reflective (A) and a formative (with causal indicators) measurement model (B), respectively (cf. [1]). In (A), the latent variable fatigue causes scores on the three items, all measuring the same construct. Item-level error is depicted by the ε′s and factor loadings by the λ′s. In (B), the not directly observable variable SES is caused by scores on the items, each measuring a different aspect of SES (i.e., income, education level, and occupation). Weights on formative construct are depicted by the γ′s, possible correlations by the r's, and the error term of the construct by the ζ. In both figures, square boxes represent observed variables, whereas the ovals represent unobserved (latent or indirectly observed) variables as estimated from the items.

having precisely the opposite causal directionality. That is, formative (as opposed to underlying) constructs are caused by scores on the observed items [2]. For example, person A is considered to have a high SES because (s)he has a high income, a high education level, and a prestigious occupation. Contrary to the assumption of unidimensionality in reflective models, indicators in formative models measure different aspects that together form a hypothesized formative factor. Consequently, indicators can, but are not required to, be correlated. Therefore, internal consistency is not an issue, but measurement error cannot be estimated at the indicator level either. Moreover, if one of the formative indicators is removed from the measurement model, the meaning of the construct the indicators form can change [2]. For example, if the income item is removed from the SES measurement model, the education and occupation items combine into a different construct that can no longer be referred to as SES (Fig. 1B). In addition, where reflective indicators are all of the same type, formative indicators can be differentiated as being of the causal, composite, or covariate indicator type [5]. Causal indicators literally cause the construct they indicate and, although unlike reflective indicators, they can cover conceptually distinct aspects of a construct, they share conceptual unity as all indicators cause the same construct. Consequently, the focal construct is considered to exist independently of its measurement and is captured as a combination of its indicators and a disturbance term [5,9,10]. Contrastingly, composite indicators are weighted and summed to create a construct without a disturbance term and do not necessarily share conceptual unity [5,10]. The focal construct is thus no latent variable that exists independently of its measurement but can be an arbitrary combination of composite indicators (e.g., an index variable) [5,11]. Finally, formative indicators

of the covariate type do not indicate the construct itself but should be included as control variables to enable adequate measurement of the focal construct [5].

The theoretical distinction between reflective and formative measurement has five major implications and consequences. First, for a complete conceptual understanding of a construct, awareness of the nature of its measurement model is crucial [12–14]. That is, a reflective measurement model is only appropriate when the construct under study corresponds to a real latent property or process. Contrastingly, a formative measurement model implies that the construct under study is an operationalization of a multidimensional [10] (social) construct (cf. [15]) or a summary of various conceptually distinct variables. Second, as shown by Law and Wong [16] and as discussed by Diamantopoulos, Riefler, and Roth [1], misspecification of a reflective measurement model as formative (or vice versa) can greatly bias estimates of structural relationships among variables and produce theoretically meaningless indices of model fit. Third, where reflective models assume unidimensionality and construct validity can be assessed through factor analysis, formative measurement models do not assume unidimensionality and their validity assessment is far more complicated (e.g., [2,9]). Although it is technically possible to estimate a statistic representing the internal consistency of a formative construct (e.g., Cronbach's alpha), the resulting estimate has no interpretational value, nor does it "prove" the existence of a single latent variable underlying the indicators (cf. [13]). As such, researchers aiming to validate formative constructs are restricted to complicated ways of assessing content (nomological), criterion (concurrent or predictive), and structural validity (e.g., [2,9]). Fourth, because a standalone construct with a formative measurement model lacks scaling, it is impossible to estimate such a model (e.g., [4]). Instead, formative models depend on the inclusion of reflective indicators/outcomes to achieve model identification [1,3]. Fifth, interventions aiming to improve the scores on a formative construct can target individual indicators of the construct as they "cause" the construct [10]. Considering these important implications, the distinction between reflective and formative measurement is not merely conceptual nitpicking. Instead, a good match between definition and measurement model specification is indispensable.

Despite its relevance, the issue of explicitly specifying a measurement model as reflective or formative is often overlooked. That is, constructs are easily assumed to have reflective measurement models, for which clear standards for validity assessment and modeling exist. For example, Jarvis et al. [2] show in an extensive review that in marketing and consumer research 336 of 365 constructs that should be modeled as formative were modeled as reflective. But the issue is likely much broader, also inflicting health measurement studies. For example, two recent studies in this journal that despite using the COSMIN checklist—a standard tool for evaluating the quality of measurement instruments [17,18]—do not explicitly mention the formative-reflective measurement distinction [19,20]. In our view, this may be due to the little attention many researchers, health measurement instruments, and guidelines such as the COSMIN checklist pay to the distinction. That is, the COSMIN only includes one item addressing the distinction and provides no further guidelines for formative measures [21]. The little attention the assessment and evaluation of formative measurement models receive might be attributed to the higher complexity of identifying and handling these models. This higher complexity has led several researchers to argue against using formative measurement [4,15,22–26]. However, although reflective measurement is preferable whenever possible, it is simply not suitable for some constructs [10,27]. Therefore, rather than abandoning formative measurement altogether, it seems instrumental to broaden researchers' perspectives and make handling formative measurement more accessible.

The present article aims to make handling the issue of reflective vs. formative measurement less daunting. Specifically, we first aim to introduce an adapted version of a systematic checklist for identifying the—reflective and/or formative—nature of a construct's measurement model based on Jarvis et al. [2]. Here, our approach consists of applying the checklist to the relevant and relatively novel construct of sustainable employability. Second, we aim to demonstrate the applicability of structural equation modeling (SEM) for handling formative constructs. To this end, we estimate a series of complementary formative models for sustainable employability that provide insight into mainly its criterion validity and the structural validity of sustainable employability as a formative construct. Finally, by subjecting sustainable employability to these conceptual and statistical analyses, we intend to provide important insights into the construct itself, the implications of its theoretical measurement model, and the weights of its components (cf. [28]). As such, "handling" in the title of this article covers identifying a construct's measurement model given its definition, validating it, and carefully considering its implications.

1.1. Sustainable employability

Sustainable employability is a relatively novel yet relevant topic in the field of occupational health research. Put simply, the construct of sustainable employability intends to capture an individual's ability to function at work and on the labor market throughout their working life (cf. [29]). This construct's relevance seems obvious as the aging of the population in most industrialized countries pushes for extended working lives. However, so goes the reasoning, working lives cannot simply be extended without any facilitation. As such, the concept of sustainable employability emerged as a relevant criterion to identify the conditions that enable individuals' long-term participation in work. However, being a novel concept, clarity on sustainable employability's conceptualization and measurement is still needed [30]. Thus, in addition to providing a generalizable illustration of handling the reflective-formative measurement

distinction, the present article also provides necessary conceptual clarity on sustainable employability.

Throughout our article, we use Fleuren et al. [31]'s definition of sustainable employability as it meets the criteria an adequate definition of sustainable employability should meet (cf. [30]):

"Sustainable employability means that an individual's ability to function in current and future work is not negatively affected by that individual's employment over time. An individual's ability to function in current and future work in this context consists of a set of characteristics (i.e., perceived health status, work ability, need for recovery, fatigue, job satisfaction, motivation, perceived employability, skill-gap, and job performance) that collectively describe the degree to which an individual can be employed at different points throughout the working life" [31].

This definition describes sustainable employability as a second-order construct with an inherently longitudinal nature. At one point in time, this second-order construct is a combination of nine first-order factors (i.e., performance, skill gap, etc.; Fig. 2) that serve as sustainable employability's indicators. These first-order factors are in turn measured by reflective items from validated scales. As such, the subject of discussion is whether the first-order factors are either reflective or formative indicators of sustainable employability. Consequently, if sustainable employability is identified as a formative construct, it would be a "reflective first-order, formative second-order construct" [2]. This is favorable as a purely formative measurement model would wrongly assume that the indicators measure the intended construct perfectly [4]. That is, self-report items are (almost) never free of measurement error and an adequate measurement model should quantify and filter item-level error out of the factors. As a reflective first-order formative second-order model allows for the estimation of item-level error [4], and the hypothetical formative measurement model for sustainable employability (See Fig. 2) would follow this structure, it would not wrongly assume absence of measurement error on the item level. As such, by including latent factors as formative indicators, this type of model differs from the exclusively formative models as typically discussed in the literature in a desirable way.

2. Methods

2.1. The checklist

To address our first aim of identifying the appropriate measurement model for sustainable employability, an adapted version of the checklist as formulated by Jarvis et al. [2] was used. This checklist consisted of a set of six objective criteria to determine whether a construct has a reflective or formative measurement model (see the left column of Table 1 for the six criteria as adapted from the original checklist). The original checklist was adjusted to achieve more parsimonious wording of the items and to incorporate the distinction between causal and composite formative indicators (item 3) [5], which was not incorporated in the original checklist.

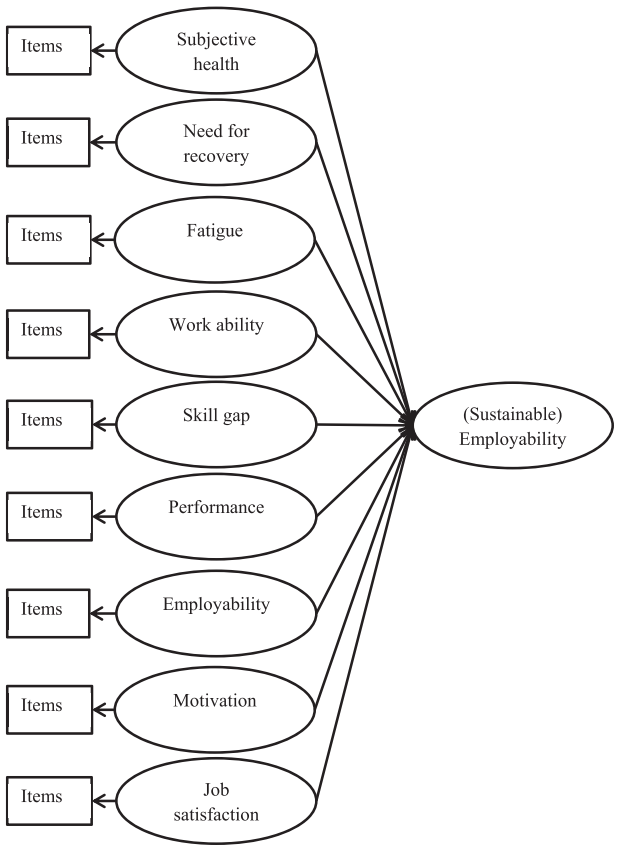

Fig. 2. Schematic depiction of the theoretical formative measurement model of sustainable employability at a single point in time.

2.2. Analyses

To address our second aim, four complementary model types were fitted to establish the structural validity of sustainable employability as formative construct. We describe the rationale and results of the most complex and notable model type in this article. Specifics regarding the other three model types can be found in our Electronic Supplement (Sections D−G, pages 17−29). Importantly, each model, including the one discussed in this article, has its own (dis)advantages and we recommend using all four modeling approaches complementarily. However, due to the extensive descriptions they require, we could not report all models in this article. All models were estimated using Mplus 7.

The main model we estimated to assess the validity of sustainable employability's formative indicators together as a set was a multiple indicator multiple cause (MIMIC) model (Fig. 3). A MIMIC model was used because standalone formative measurement models would not be identified using SEM software [5]. Therefore, we used three items capturing perceived ability and willingness to work until the age eligible for retirement and to achieve model identification.

Table 1. Application of Jarvis et al.'s [1] checklist to sustainable employability (SE)

Checklist item	Application to sustainable employability
1. Are the indicators (items) (A) defining characteristics or (B) manifestations of the construct?—"A" indicates a formative and "B" a reflective measurement model.	As mentioned in the article, the indicators of SE are the first-order factors (i.e., performance, skill gap, employability, need for recovery, fatigue, subjective health, workability, motivation, and job satisfaction). The definition states "an individual's employability in this context consists of a set of characteristics … that describe the degree to which an individual can be employed in a type of employment situation". The word "describes" indicates that the first-order factors are defining characteristics of SE. Therefore, the answer to this question is A suggesting formative measurement.
2. Would changes in the indicators/items cause changes in the construct or the other way around?—The former indicates formative and the latter reflective.	For SE, it seems counter-intuitive that the construct itself would change, independently of a change in its indicators. Rather, a specific aspect of the construct would change (i.e., one of the first-order indicators), and as a result the overall construct would change. For example, an individual's health may be affected negatively, which reduces that individual's ability to function in work. Thus, causality flows from the first-order factors to the second-order factor and a formative measurement model would be more appropriate.
3. Should each indicator capture exactly the same? "Yes" indicates reflective; "no, but they share conceptual unity in terms of causing a common construct" indicates causal formative; and "not at all" indicates composite formative indicators.	The first-order factors in the definition of SE are drawn from three different disciplines and are all selected to cover a different aspect of SE. Therefore, each first-order indicator is different from the others and a formative measurement model would be more appropriate. Furthermore, SE's formative indicators should theoretically all be distinct aspects of SE that all provide a theoretically meaningful contribution. In that sense, they do share conceptual unity in this sense and are not an arbitrary mix of indicators. Therefore, SE's formative indicators are of the causal rather than the composite kind.
4. Would dropping one of the indicators alter the conceptual domain of the construct? "Yes" indicates formative; "no" indicates reflective.	Each of the first-order indicators of SE covers a distinct aspect. If one of the indicators would be dropped, the second-order construct SE would no longer include that aspect and would be conceptually different from the construct with the indicator included. Again, a formative measurement model seems more appropriate.
5. Should a change in one of the indicators be associated with changes in the other indicators? "Yes" indicates reflective; "no" indicates formative.	The indicators of SE are potentially but not necessarily interrelated. For example, a change in health can be related to a change in skill gap; someone may lose his arms and his typing skills simultaneously. However, this is not always the case as someone may suffer from an illness but can still be fully skilled for the job. Moreover, a change in need for recovery may only manifest itself in a change in work ability at a later point in time. Therefore, changes in one of SE's first-order indicators can but do not need to be related to changes in the others. Thus, a formative measurement model seems more appropriate.
6. Are the indicators expected to have the same antecedents and consequences? "Yes" indicates reflective; "no" indicates formative.	The indicators of SE can be affected by different aspects of employment. For example, working time arrangements can cause increases in need for recovery and fatigue, whereas skill gap and job satisfaction may remain unaffected. Thus, also for this final question the answer indicates that a formative measurement model should be used to conceptualize SE.

These items were selected because the reflective indicator(s) in a MIMIC model should either theoretically be outcomes of the construct or capture it entirely. As the perceived ability and willingness to work items were used in a previous study to operationalize sustainable employability [32] and should theoretically be outcomes of sustainable employability (i.e., as reality precedes perception), we considered them adequate reflective indicators for our model. In the resulting MIMIC model, the formative construct as scaled by the perceived ability and willingness to work items was regressed on the formative indicators (i.e., the nine dimensions of sustainable employability) to obtain weights for each of the formative indicators. Indices of model fit [1,3] interpreted using conventional criteria (i.e., Comparative fit index [CFI] > 0.90, Tucker-Lewis index [TLI] > 0.90, root mean square error of approximation [RMSEA] < 0.05) [33,34]

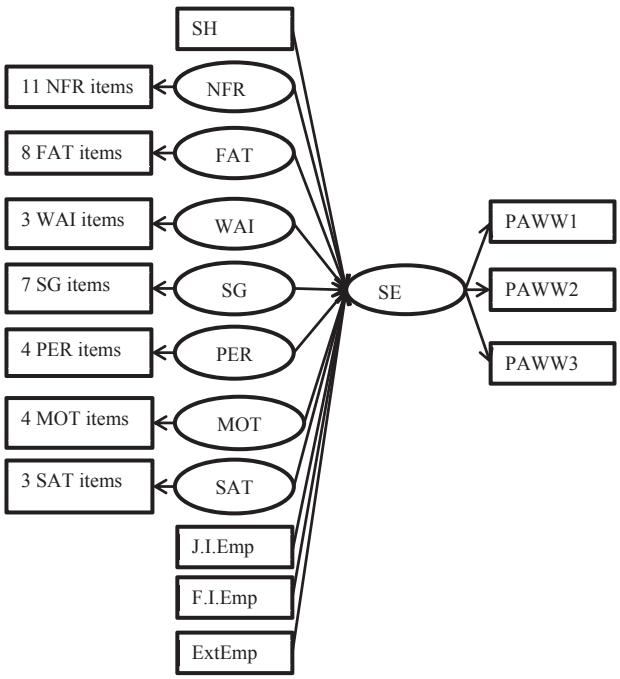

Fig. 3. Schematic depiction of the MIMIC model including all of the formative indicators of SE as and the three PAWW items as reflective indicators. The model depicts all formative indicators of SE as well as the three reflective PAWW items. Square boxes represent observed variables, whereas the ovals represent unobserved (latent or indirectly observed) variables as estimated from the items. All variables in the middle column are the formative indicators of SE. All variables in the right column are PAWW items. All variables in the left column are items corresponding to the latent formative indicators of SE. ExtEmp, external employability; FAT, fatigue; F.I.Emp, firm internal employability; J.I.Emp, job internal employability; MOT, motivation; NFR, need for recovery; PAWW, perceived ability and willingness to work until the official retirement age; PER, performance; SAT, job satisfaction; SE, sustainable employability; SG, skill gap; SH, subjective health; WAI, work ability.

and the variance of the error term of the construct [1,35,36] was used to assess structural validity.

2.3. Sample

The aforementioned analyses were performed using data from the prospective Maastricht Cohort Study. The ongoing Maastricht Cohort Study started in May 1998 as a large scale longitudinal study on various relationships between aspects of work and the development of fatigue. At baseline (i.e., 1998), it included 12,140 Dutch respondents, aged between 18 and 65 years, working in 45 different organizations. All respondents provided written informed consent and data for all of the Maastricht Cohort Study waves were collected in accordance with the Declaration of Helsinki (1964) and its later amendments (see [37] for more information). The protocol for the Maastricht Cohort Study is approved by the Medical Research and Ethics Committee of the Maastricht University Medical Center under number MEC 08-4-032.

Specifically, data from the 2012 measurement wave of the Maastricht Cohort Study were used. This measurement wave contained all of the self-report measures necessary to analyze the MIMIC model. Before exclusion, this data set contained 4,783 respondents (1207 female, 3,497 male, 85 missing), ranging in age from 34.59 to 78.97 years old (M = 57.57, SD = 8.43, n = 4,783). As measures used in the analyses were designed for working individuals, a total of 2,239 respondents were excluded because they were not working at time of survey completion or because they had more than one job (details regarding exclusion criteria can be found in section A, page 2 of the Electronic Supplement). A total of 2,544 respondents (771 female, 1,736 male, 37 missing) ranging in age from 34.59 to 71.33 (M = 53.14, SD = 6.34) remained for inclusion in the analyses.

2.4. Instruments

The Maastricht Cohort Study 2012 wave contained many self-report measures for various work, work-context, and individual characteristics. Among this set were subjective health, need for recovery, fatigue, work ability, skill gap, performance, employability, motivation to work, and job satisfaction, and perceived ability and willingness to work until the age eligible for retirement. Brief descriptions of the scales that were used in the analyses can be found in Table 2. Further descriptions of the scales as well as their factor structure can be found in the Electronic Supplement (section B, pages 3−6).

3. Results

3.1. Application of the checklist to sustainable employability

As demonstrated in Table 1, applying the checklist reveals that sustainable employability should be considered as a formative construct, necessitating a measurement model suited for such a construct. Specifically, as the relationship between the observed indicators (i.e., scale items) and the first-order factors is reflective, a so-called "reflective first-order, formative second-order model" [2] most adequately describes sustainable employability (Fig. 2). Sustainable employability is thus a formative combination of nine reflective first-order factors. Moreover, considering the response to item three, sustainable employability's formative indicators seem to be of the causal type specifically.

3.2. Modeling sustainable employability as a formative construct

In this section, we discuss the results of the main MIMIC model as described under "Analyses". Results from the other complementary models and confirmatory factor analyses can be found in the Electronic Supplement. As described in the Supplement (sections C-H, pages 7−32), all models fitted the data well and produced expected results.

Table 2. Summary of scales used in analyses

Construct measured	Scale used	Sample item
1. Subjective health	Single item; MOS 36-item Short-Form Health Survey (SF-36) questionnaire [38].	''In general, would you say your health is 1) excellent; 2) very good; 3) good; 4) fair; 5) poor''
2. Need for recovery	11 items; Need for recovery scale from the Dutch Questionnaire on the Experience and Evaluation of Work (VBBA) [39,40].	''By the end of the working day, I feel really worn out''; Yes/No
3. Fatigue	Eight items; Subjective experience of fatigue subscale from the Checklist Individual Strength (CIS) [41].	''I feel tired''; seven-point Likert
4. Work ability	Three items; Work Ability Index [42,43].	''If you were to rate your work ability as a 10 out of 10 in the best period of your life, how would you rate your current work ability on a 10 point scale?''
5. Skill gap	Nine items; adapted from the HBO-monitor [44].	''How would you rate your own level of Skill X''; five-point low-high. ''What is the level of Skill X required by your job''; five-point low-high
6. Employability	Three single items; one self-constructed, two adapted from a self-report employability questionnaire by De Cuyper and De Witte [45]. Each measured a different aspect of employability	''I am convinced that I could keep my current job until retirement, if I wanted to''; five-point Likert
7. Performance	Four items; adapted from the Dyne and LePine [46] Core Task Performance Scale.	''I meet the performance standards of my job''; five-point Likert
8. Motivation	Four items; Motivation subscale from the CIS [41].	''I feel no desire to do anything''; seven-point Likert
9. Job satisfaction	Three items; from a shortened 12-item version of the Utrecht Work Engagement Scale (UWES) [47].	''I am enthusiastic about my job''; seven-point never-every day
10. Perceived ability and willingness to work until the official retirement age (PAWW)	Three items; adapted versions of items from van Dam, van der Vorst and van der Heijden [48].	''I believe to be mentally able to continue working in my current job until the retirement age that currently applies to me''; five-point Likert

The main MIMIC model we tested included sustainable employability's formative indicators and the three perceived ability and willingness to work items as reflective indicators (Fig. 3). The MIMIC model fitted the data well (χ^2 = 3,670.051, df = 1073, $P > 0.01$, CFI = 0.937, TLI = 0.930, RMSEA = 0.031 [90% C.I. = 0.030–0.032]) and six of 11 path coefficients between the sustainable employability's formative indicators and the second-order factor sustainable employability were significant (Table 3). Moreover, the R-square of the second-order factor sustainable employability was 0.30, indicating the amount of variance in sustainable employability as explained by its formative indicators. These findings suggested that a formative measurement model for sustainable employability fitted the data well and that the formative indicators functioned well together as a set.

4. Discussion

This article intends to provide researchers with tools to handle the reflective-formative measurement conundrum. To that end it first demonstrates how the nature of a construct's measurement model can be identified using an adapted version of the checklist by Jarvis et al. [2]. Second, it shows how SEM can be used to gain insight into the criterion and structural validity of modeling formative indicators to form a single construct. By using sustainable employability (sustainable employability; [31]) as a working example, the article also provides much needed clarity on this construct's nature and validity as a formative construct. Specifically, we show it is feasible to measure sustainable employability as a ''reflective first-order, formative second-order construct'' [1], where the formative indicators are subjective health, need for recovery, fatigue, work ability, performance, employability, skill gap, motivation, and job satisfaction. In addition, the results from the models in this article and the supplement may inform researchers about possible weights for sustainable employability's indicators in future studies and may hint to the relative importance of each component. Although our approach has some limitations (as discussed below), it does provide researchers with the means to handle formative construct as well as an elaboration of the extensive implications

Table 3. Path coefficients for the effects of first-order factors on the second-order factor employability consisting of three reflective indicator items in a full MIMIC model

First-order factor	Path coefficient	Standard error	Est./S.E.	*P*-value
1. Subjective health (one item)	0.028	0.025	1.103	0.270
2. Need for recovery	0.211[a]	0.035	5.975	0.001
3. Fatigue	0.053	0.045	1.178	0.239
4. Work ability	0.343[a]	0.040	8.500	0.001
5. Skill gap	0.031	0.024	1.290	0.197
6. Performance	−0.029	0.023	−1.257	0.209
7. Employability (three separate items)				
1) Job internal employability	0.141[a]	0.019	7.626	0.001
2) Firm internal employability	0.024	0.020	1.236	0.216
3) External employability	−0.060[a]	0.020	−2.965	0.003
8. Motivation	−0.123[a]	0.042	−2.929	0.003
9. Job satisfaction	0.058[a]	0.026	2.255	0.024

Note. Reported path coefficients are standardized, transcribed from the Mplus 7 STDYX output.
[a] Significant at $P < 0.05$.

that come with it (see further in this section). As such, researchers working with complex (health) constructs such as quality of life, functioning, and severity of disease can use this article to check assumptions about the measurement models of these constructs and to model them more appropriately if they are formative. Consequently, research and interventions involving such complex constructs can be performed more adequately, so that biased estimations and ineffective approaches can be detected and avoided. In some cases, this may require reconsideration of constructs that are currently perceived as valid and used in practice.

4.1. Limitations

A first limitation to this study is that the results of the MIMIC models we estimate depend on the choice of the reflective indicators for sustainable employability. That is, the reflective indicators determine the path coefficients of the formative indicators to the construct as well as their significance levels. However, as no alternative approaches to assessing the structural validity of formative constructs exist, this limitation could not be avoided. Therefore, we carefully selected straightforward outcomes of sustainable employability as reflective indicators in our models (i.e., perceived ability and willingness to work until the retirement age). Moreover, we tested several complementary models of which some did not include the additional perceived ability and willingness to work items as reflective indicators (see Supplement Sections G and H). As such, the indicators we use should theoretically be appropriate and it is unlikely that our results are biased because of this limitation. In addition, this limitation would only apply to our models for sustainable employability and not to the general illustration provided in this article. Nonetheless, this point does underscore the importance of carefully selecting appropriate reflective indicators for MIMIC models.

A second limitation specific to our example of sustainable employability is that we only investigate its measurement model at one point in time. This is suboptimal as sustainable employability is an inherently longitudinal construct [31] but could not be avoided due to lack of longitudinal data for most of the formative indicators used (except for those in the health domain of sustainable employability). Nonetheless, assessing sustainable employability's validity as a formative construct at one point in time is challenging enough and provides a basis for future studies that do take this longitudinal perspective. Specifically, such future studies could particularly consider sustainable employability's criterion validity more extensively by predicting employment status on the long term and timing of retirement, potentially across different groups.

Third, some of sustainable employability's formative indicators may not have been measured in the best possible way. That is, we used an existing data set (i.e., the 2012 wave of the Maastricht Cohort Study) that featured all of the necessary constructs. As such, concessions had to be made, resulting in some of the variables to be captured by suboptimal measures (in particular employability, job satisfaction, and motivation). Although factor analyses suggested that the measures did function appropriately, future research should aim to include optimal scales for all constructs. In addition, the models we tested show acceptable fit even with suboptimal measures, which might arguably only show the robustness of the approach.

A fourth limitation relates to the retention of the formative indicators of sustainable employability. That is, ideally all path coefficients in the MIMIC model would have been significant. However, our findings indicate otherwise. These findings are probably due to interrelatedness among sustainable employability's formative indicators. For example,

if work ability and subjective health are interrelated they could partially explain the same variance in the sustain employability construct, rendering either or both indicator's path coefficients insignificant. As insignificance could signal redundancy [49], it may be intuitive to omit all formative indicators with insignificant path coefficients in the MIMIC model. However, this is not necessarily the best approach because these indicators may still be theoretically relevant. Moreover, as shown in the Electronic Supplement, all indicators do demonstrate criterion validity in complementary models. Therefore, this final limitation underscores the relevance of using several complementary models when handling formative constructs.

Finally, a question of completeness remains. That is, for formative constructs, there is no way of knowing whether the intended or described construct is complete, given the indicators included in its measurement model (cf. [50]). This issue is unique to constructs with formative indicators, as reflective indicators are interchangeable and should all capture the same single construct. When considering completeness, the only criterion is that the measurement model as specified exactly matches the definition of the intended construct (i.e., content validity). Therefore, although researchers working with novel formative constructs may desire full confidence that their construct encompasses the intended construct completely, this cannot be achieved with statistical tests. Instead, completeness and content validity are about the match between theory and operationalization/measurement. Based on theoretical considerations and the definition of sustainable employability in Fleuren et al. [31], the formative indicators in our model seem to cover sustainable employability adequately.

4.2. Implications

First, the application of the checklist by Jarvis et al. [2] as demonstrated in this article gives researchers an easy-to-use tool to identify the nature of a constructs measurement model. The six checklist items can be answered with relative ease to identify a construct as formative and/or reflective. Notably, as evidenced by this checklist, the reflective-formative distinction is theoretical and thus should not be made on statistical grounds (e.g., comparing models is inappropriate and impossible). However, as discussed by Bollen and Ting [51], researchers can use a tetrad test when they remain in doubt whether their construct of interest is formative (causal) or reflective. Still, as a construct's measurement model depends on the construct's definition, it may remain topic of debate. In such cases, the checklist as it may help to structure theoretical debates and to reach a well-grounded conclusion. For example, careful application of the checklist to complex constructs such as quality of life, functioning, severity of disease, and psychological disorders may provide clarity on and deepen our understanding of these constructs. Importantly, this implies that researchers may have to reconsider constructs they have been using in studies for years. Moreover, developers of practical interventions involving such constructs may have to go back to the drawing board, and priorities in both research and practice may need revision as well.

A second important consequence of this article is that it demonstrates SEM as a useful tool to handle formative constructs. That is, by using sustainable employability as an example in several complementary models, we demonstrate that SEM can be used to gain insight into the structural and criterion validity of complex formative constructs. This particularly applies to formative constructs with causal rather than composite indicators, but as Edwards and Bagozzi [52] argue, claiming a construct is a composite variable without exploring its validity is a bad practice. Importantly, as discussed, there certainly are some pitfalls to be aware of (i.e., issues of completeness [50], collinearity among formative indicators [53], and item-level error and model identification [4]) when using presented modeling strategies, but these strategies are still superior to standard approaches for reflective measurement models (e.g., factor analyses and Cronbach's alpha). After all, these latter approaches have exactly no meaning for constructs that include formative indicators and their application to formative constructs should therefore be eradicated. With the demonstration, this article provides researchers with the tools to do so.

Third, if anything, this article has illustrated that formative constructs deserve more attention considering their complexity and extensive implications. As such, we think that tools for assessing (health) measurement instruments should include more elaborate directions on handling formative measurement models. For example, in its current format, the COSMIN checklist only includes a single item on formative vs. reflective measurement [18,21]. This does not do justice to—nor helps researchers struggling with—the complexity of formative measurement. Given the applicability of Jarvis et al.'s checklist combined with SEM to explore formative constructs in detail as demonstrated in this article, they could be useful extensions of checklists for assessing the quality of measurement instruments.

Finally, our findings have important implications for the use of sustainable employability in interventions and research. First, recognizing sustainable employability as a formative construct implies that interventions aiming to improve sustainable employability can theoretically target its formative indicators. That is, scores on formative constructs are determined by scores on their indicators (and not inversely as for reflective constructs). As such, interventions aiming to promote sustainable employability can theoretically target its individual constituents. Importantly, however, increases in one formative indicator of sustainable employability could be compensated by decreases in another (which would not apply to reflective indicators). Moreover, sustainable employability as a whole is what HR practitioners strive to improve and what occupational health researchers aim to predict and understand.

Therefore, sustainable employability's indicators should be considered both separately (as targets in interventions) and jointly (for scientific understanding and societal relevance) (e.g., [54]). Second, when considering formative indicators as separate entities, it theoretically makes sense to model relationships among indicators in, for example, network models (cf. [55]). Considering sustainable employability's inherent connection with time, it may then be attractive to explore causality among its formative indicators at different time points (e.g., skill gap on t_0 might predict need for recovery on t_1). These implications could lead to a better understanding of and interventions for sustainable employability specifically, but similar approaches might apply to other (health) constructs.

4.3. Conclusion

In conclusion, this article contributes to the literature by providing an illustration of how relatively accessible tools for identifying (i.e., a checklist) and handling (i.e., SEM) formative constructs can be used. These tools enable a more adequate handling of formative constructs, so that well-founded conclusions regarding structural relations between such constructs and others can be drawn. As argued, the formative-reflective measurement distinction may thus require us to reconsider some of our well-established (health) constructs (e.g., quality of life, severity of disease, and functioning). In the end, if we wish to improve the quality of our lives, functioning and health, a thorough understanding of such constructs is indispensable.

Acknowledgments

The authors are grateful to Jos Slangen from Maastricht University for his excellent assistance regarding data management. The first author is grateful to Stefan Gruijters and Huub Hoofs, both from Maastricht University, for inspiring side-discussions on psychometrics that have sharpened the thinking underlying this article.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2018.07.007.

References

[1] Diamantopoulos A, Riefler P, Roth KP. Advancing formative measurement models. J Bus Res 2008;61:1203−18.
[2] Jarvis CB, MacKenzie SB, Podsakoff PM. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J Consum Res 2003;30(2):199−218.
[3] Diamantopoulos A, Winklhofer HM. Index construction with formative indicators: an alternative to scale development. J Marketing Res 2001;38(2):269−77.
[4] Edwards JR. The fallacy of formative measurement. Organ Res Method 2011;14(2):370−88.
[5] Bollen KA, Bauldry S. Three Cs in measurement models: causal indicators, composite indicators, and covariates. Psychol Methods 2011;16(3):265−84.
[6] Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. J Roy Stat Soc 2002;165(2):233−61.
[7] Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. J Pers Assess 2003;80(3):217−22.
[8] Bollen KA, Lennox R. Conventional wisdom on measurement: a structural equation perspective. Psychol Bull 1991;110(2):305−14.
[9] Hardin AM. A call for theory to support the use of causal-formative indicators: a commentary on Bollen and Diamantopoulos (2017). Psychol Methods 2017;22(3):597−604.
[10] Bollen KA, Diamantopoulos A. In defense of causal-formative indicators: a minority report. Psychol Methods 2017;22(3):581−96.
[11] Bollen KA. Interpretational confounding is due to misspecification, not to type of indicator: comment on Howell, Breivik, and Wilcox (2007). Psychol Methods 2007;12(2):219−28.
[12] Borsboom D, Mellenbergh GJ, van Heerden J. The theoretical status of latent variables. Psychol Rev 2003;110(2):203−19.
[13] Gruijters SLK, Fleuren BPI. Measuring the unmeasurable: the psychometrics of life-history strategy. Hum Nat 2018;29(1):33−44.
[14] Borsboom D, Mellenbergh GJ, van Heerden J. The concept of validity. Psychol Rev 2004;111(4):1061−71.
[15] Guyon H. The fallacy of the theoretical meaning of formative constructs. Front Psychol 2018;9:179.
[16] Law KS, Wong C. Multidimensional constructs in structural equation analysis: an illustration using the job perception and job satisfaction constructs. J Manage 1999;25(2):143−60.
[17] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34−42.
[18] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratfort PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539−49.
[19] Wong CKH, Lang BHH, Lam CLK. A systematic review of quality of thyroid-specific health-related quality-of-life instruments recommends ThyPRO for patients with benign thyroid diseases. J Clin Epidemiol 2016;78:63−72.
[20] van Melle MA, van Stel HF, Poldervaart JM, de Wit NJ, Zwart DLM. Validation of a questionnaire measuring transitional patient safety climate indicated differences in transitional patient safety climate between primary and secondary care. J Clin Epidemiol 2018;94:114−21.
[21] Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2011;21:651−7.
[22] Bagozzi RP. On the meaning of formative measurement and how it differs from reflective measurement: comment on Howell, Breivik, and Wilcox (2007). Psychol Methods 2007;12(2):229−37.
[23] Howell RD. What is the latent variable in causal indicator models? Measurement 2014;12(4):141−5.
[24] Howell RD, Breivik E, Wilcox JB. Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007). Psychol Methods 2007;12(2):238−45.
[25] Howell RD, Breivik E, Wilcox JB. Reconsidering formative measurement. Psychol Methods 2007;12(2):205−18.

[26] Wilcox JB, Howell RD, Breivik E. Question about formative measurement. Psychol Methods 2008;61(12):1219−28.
[27] Bollen KA, Diamantopoulos A. Notes on measurement theory for causal-formative indicators: a reply to Hardin. Psychol Methods 2017;22(3):605−8.
[28] Hardin AM, Chang JCJ, Fuller MA, Torkzadeh G. Formative measurement and academic research: in search of measurement theory. Educ Psychol Meas 2011;71(2):281−305.
[29] van der Klink JJL, Bültmann U, Burdorf A, Schaufeli WB, Zijlstra FRH, Abma FI, et al. Sustainable employability − definition, conceptualization, and implications: a perspective based on the capability approach. Scand J Work Environ Health 2016; 42(1):71−9.
[30] Fleuren BPI, de Grip A, Jansen NWH, Kant I, Zijlstra FRH. Critical reflections on the currently leading definition of sustainable employability. Scand J Work Environ Health 2016;42(6):557−60.
[31] Fleuren BPI, de Grip A, Jansen NWH, Kant I, Zijlstra FRH. Sustainable employability: a multidisciplinary approach 2016: Securex' International Conference on Sustainable Employability 2016 in Brussels and at APA and NIOSH's Work, Stress, and Health Conference 2015 in Atlanta.
[32] Oude Hengel KM, Blatter BM, Geuskens GA, Koppes LLJ, Bongers PM. Factors associated with the ability and willingness to continue working until the age of 65 in construction workers. Int Arch Occup Environ Health 2012;85(7):783−90.
[33] Hu LT, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999;6(1):1−55.
[34] MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods 1996;1(2):130−49.
[35] Diamantopoulos A, Siguaw JA. Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration. Br J Manage 2006;17(4):263−82.
[36] Williams LJ, Edwards JR, Vandenberg RJ. Recent advances in causal modeling methods for organizational and management research. J Manage 2003;29(6):903−36.
[37] Kant I, Bultmann U, Schroer KAP, Beurskens AJHM, van Amelsvoort LGPM, Swaen GMH. An epidemiological approach to study fatigue in the working population: the Maastricht cohort study. Occup Environ Med 2003;60:32−9.
[38] Ware JE, Sherbourne CD. The mos 36-item short-form health survey (Sf-36): conceptual-framework and item selection. Med Care 1992;30:473−83.
[39] van Veldhoven M, Broersen S. Measurement quality and validity of the "need for recovery scale". Occup Environ Med 2003;60(Suppl 1):i3−9.
[40] van Veldhoven M, Meijman TF. Het meten van psychosociale arbeidsbelasting met een vragenlijst: de vragenlijst beleving en beoordeling van de arbeid (VBBA) [The measurement of psychosocial job demands with a questionnaire (VBBA)]. Amsterdam: NIA; 1994.
[41] Vercoulen JHMM, Swanink CMA, Fennis JFM, Galama JMD, Vandermeer JWM, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994;38(5):383−92.
[42] Ilmarinen J. The work ability index (WAI). Occup Med 2007;57(2):160.
[43] Ilmarinen J. Towards a longer worklife: aging and the quality of worklife in the European Union. Helsinki: National Institute for Occupational Health; 2006.
[44] Humburg M, van der Velden R. What is expected of higher education graduates in the 21st century?. In: Buchanan J, Finegold D, Mayhew K, Warhurst C, editors. Oxford Handbook of Skills and Training. Oxford: Oxford University Press; 2016.
[45] de Cuyper N, de Witte H. The management paradox: self-rated employability and organizational commitment and performance. Pers Rev 2011;40(2):152−72.
[46] Dyne LV, LePine JA. Helping and voice extra-role behaviors: evidence of construct and predictive validity. Acad Manage J 1998; 41(1):108−19.
[47] Schaufeli WB, Salanova M, Gonzalez-Roma V, Bakker AB. The measurement of engagement and burnout: a confirmative analytic approach. J Happiness Stud 2002;3:71−92.
[48] van Dam K, van der Vorst JDM, van der Heijden BIJM. Employees' intentions to retire early: a case of planned behavior and anticipated work conditions. J Career Dev 2009;35(3):265−89.
[49] Costa DSJ. Reflective, causal, and composite indicators of quality of life: a conceptual or an empirical distinction? Qual Life Res 2015; 25:2057−65.
[50] Ford LA. Selection issues of formative models. J Manage Dev 2017; 36(5):660−70.
[51] Bollen KA, Ting K. A tetrad test for causal indicators. Psychol Methods 2000;5(1):3−22.
[52] Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. Psychol Methods 2000; 5(2):155−74.
[53] Cadogan JM, Lee N. Improper use of endogenous formative variables. J Bus Res 2013;66(2):233−41.
[54] Fleuren BPI, van Amelsvoort LGPM, de Grip A, Zijlstra FRH, Kant I. Time takes us all? A two-wave observational study of age and time effects on sustainable employability. Scand J Work Environ Health 2018; 1−11.
[55] Borsboom D, Cramer AOJ. Network analysis: an integrative approach to the structure of psychopathology. Annu Rev Clin Psychol 2013;9:91−121.